C3X: Correlation, Causation and Controlled Experimentation for C2

Paul J. Hiniker, Ph.D.,
Defense Information Systems Agency (DISA/GE)
5275 Leesburg Pike
Falls Church, VA 22041
703-882-1785
 hinikerp@ncr.disa.mil

ABSTRACT

   This paper examines the key role of controlled experimentation in testing causal hypotheses on the warfighting effectiveness of C2 technologies and procedures. Through the years many hypotheses have been advanced regarding the factors making for effective warfighting. More troops, more firepower, higher speed of maneuver, superior doctrine, better training and superior C4I (Command, Control, Communications and Intelligence) are all factors hypothesized to make for more effective warfighting. Warfighting, itself, is adjudged more effective when enemy combat losses appreciably exceed own force losses. How does one go about testing the many hypotheses on the causes of warfighting effectiveness against observational evidence? For example, what difference does it make to warfighting effectiveness if we put in a particular new C2 technology? We don't need to know simply what's the difference between military units with and without the new C2 technology; we need to know what actually happens with the new C2 technology compared with what would have happened without it. We shall prove that controlled experiments provide the only unequivocal tests of such causal hypotheses; otherwise the observed results are open to rival explanation in terms of causation by some of the uncontrolled factors. We introduce causal hypothesis testing with observations on a single group and then move to the method of using simple correlational data for two groups. This forces us to confront the open ended issue of control groups and control variables in testing causal hypotheses which in turn leads us to consider the most conclusive method, controlled experimentation. We then demonstrate the feasibility and utility of this method by providing examples of substantial results from six controlled experiments on the causes of warfighting effectiveness: two on the effects of alleged superior doctrine, *viz*. use of contingency planning, and four on the effects of alleged superior C4I, *viz*. use of the Common Operational Picture and use of a prototyped planning aid. Finally, we examine some implications of this testing method for evaluating the tenets of Network Centric Warfare and associated technologies. The experimental method used and advocated here for effectiveness testing of proposed defense capabilities and technologies is essentially the same as the method of randomized clinical trials employed in the health sciences to determine whether or not the use of potential new healing drugs causes improved health.

1. REPORT DATE SEP 2004	2. REPORT TYPE	3. DATES COVERED 00-00-2004 to 00-00-2004
4. TITLE AND SUBTITLE C3X: Correlation, Causation and Controlled Experimentation for C2		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Defense Information Systems Agency,DISA/GE,5275 Leesburg Pike,Falls Church,VA,22041		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited		
13. SUPPLEMENTARY NOTES The original document contains color images.		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified		25	

Introduction

For centuries contingency planning has been recommended as a superior military practice. Does use of contingency planning by a battlefield commander, in fact, cause improved warfighting effectiveness? We could observe the use of contingency planning (x) in the operational setting of a battle or military exercise and see if the Blue commander destroyed more Red platforms than he lost (y). But even were this to occur, to claim that x caused y remains open to the obvious rival explanation that some other variable (c) or combination of variables occurring simultaneously with the battle or exercise, e.g. use of advanced weaponry, was the true cause of Blue's success. To draw such an inference from the observations is to be guilty of the classic *ad hoc propter hoc* fallacy. Only if we could isolate the cause would such a procedure provide a convincing test of our hypothesis. As an improvement on the above, we could intervene in an exercise to delay the onset of contingency planning, measure y, then insert x and remeasure y at end exercise. But even here, any improvement in y could be criticized as having arisen not from x but from other new events surrounding x that occurred simultaneously with x during the later phase of the exercise, e.g. Blue troops have learned more about Red causing them to perform better in the second phase. So while always instructive and often productive of valuable insights and new hypotheses, neither single shot case studies nor before and after measures on a single group provides a definitive test of a causal hypothesis.[1] To rid ourselves of such rival explanations for our findings, we need a comparison group to ascertain what would have happened if contingency planning were not used, i.e. we must deal with the counterfactual conditional.

Simple Correlation with Two Groups

As a test of our causal hypothesis, we could compare the combat wins (y) of a group of commanders who employed contingency planning (x) with that of another group of commanders who did not (~x). Passive observational evidence for and against the contingency planning/warfighting effectiveness hypothesis could be gleaned from the history of past battles or military exercises. We can array notional findings of research into two dozen such battles in a fourfold table as shown in Figure 1 below and calculate the correlation between the variables, Φ_{xy}. Here x is the independent variable, usually a defense capability; and y is the dependent variable, usually a warfighting effectiveness measure.

---

[1] Shadish, W., Cook, T., and Campbell, D. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference,* Boston: Houghton Mifflin, 2001.

Figure 1. Use of Contingency Planning (x) by Combat Outcome (y)

	~x	x	
y	4 (.17)	8 (.33)	12 (.50)
~y	8 (.33)	4 (.17)	12 (.50)
	12	12	24
	(.50)	(.50)	(1.00)

(1) Simple Phi Coefficient of Correlation[2]

$$\phi_{xy} \;=\; (\; Pxy \; - \; PxPy)\; /\; \sqrt{(\; Px\; Qx\; Py\; Qy)}$$

$$\Phi xy \;=\; (\; .33 \;\; - \; .50 \times .50)\; /\; \sqrt{(\; .50 \times\; .50 \times .50 \times .50)}\; =\; .32$$

ϕ_{xy} is simply a measure of association or correlation between two dichotomous variables where the numerator is the difference (.08) between the empirically observed association of x and y (.33) and what would logically be expected for their joint occurrence assuming statistical independence of x and y, where half the battles involved Blue use of contingency planning and half were Blue wins (.50 X .50 = .25). As shown in Equation (1), this degree of association is assessed relative to the denominator which measures the total variability in x and y ($\sqrt{}$ (.5 X .5 X .5 X .5) = .25). Random association would yield $\phi_{xy} = 0$. Clearly, we observe a tendency for use of contingency planning to be disproportionately associated with, i.e. correlated with, successful combat outcomes ($\phi_{xy} = .32$). Indeed, two thirds of the battles surveyed were either contingency planning wins or non-contingency planning losses. So we find that use of contingency planning is correlated with combat success, and we cannot reject our causal hypothesis with these data.

---

[2] The direct analogy to the Φ_{xy} coefficient for correlation of interval scale variables is the Pearson product moment correlation coefficient: $r_{xy} = \Sigma\, xy\, /\, N\sigma_x\sigma_y$. See Quinn McNemar, *Psychological Statistics*. NY: Wiley, 1962

The Role of Controls in Ruling Out Rival Explanations of Observations

Suppose, however, someone advances a rival hypothesis to explain our findings, suggesting that it is not the use of contingency planning, per se, but superior military training that caused the successful combat outcomes. After investigating the military background of the 24 Blue battle commanders, he is faced with the subdivided array of data shown in Figure 2 below.

Figure 2. Use of Contingency Planning by Combat Outcome Controlling for Training(c)

	c				~c		
	~x	x			~x	x	
y	0	8	8	y	4	0	4
~y	4	0	4	~ y	4	4	8
	4	8	12		8	4	12

$$(2) \text{ Covariance/Partial Correlation Theorem}[3]$$

$$\phi_{xy} = \phi_{xyc} \cdot P_c \sqrt{(P_{xc} Q_{xc} P_{yc} Q_{yc} / P_x Q_x P_y Q_y)} + \phi_{xy\sim c} \cdot P_{\sim c} \sqrt{(P_{x\sim c} Q_{x\sim c} P_{y\sim c} Q_{y\sim c} / P_x Q_x P_y Q_y)} + \phi_{yc} \phi_{cx}$$

Although we cannot completely reject our contingency planning/ warfighting effectiveness hypothesis with the new correlations, we find an equally plausible rival explanation for the findings, viz. training at U. S. Military Academy at West Point (c) leads to improved combat effectiveness (Φ_{cy} = .32), and West Point training is disproportionately associated with use of contingency planning (Φ_{cx} = .32). So, based on the simple correlations alone one could assert with equal confidence that West Point training caused the success in combat. Indeed, for all we know, it may have been the case that the winning battles were all correlated with yet another potential causal factor,

---

[3] The analogy to the conditional, within group, Φ correlation coefficient for interval scale variables is the partial correlation coefficient, $r_{xy.c} = (r_{xy} - r_{xc} r_{yc}) / \sqrt{(1 - r_{xc}^2)} \sqrt{(1 - r_{yc}^2)}$.
This partial correlation coefficient represents the correlation between two variables, x and y, when the influence of a third variable, c, has been controlled. The Covariance Theorem for dichotomous attributes states that for any two attributes, x and y, and a third "control" attribute, c, it is possible to equate the universal covariance, C_{xy} with a weighted average of covariances within control subgroups, and, in addition, a term involving a product of the covariances between y and c, and c and x:
$C_{xy} = P_c C_{xyc} + P_{\sim c} C_{xy\sim c} + C_{yc} C_{cx} / P_c P_{\sim c}$. Substituting $\Phi_{xy} \sqrt{(P_x Q_x)} \sqrt{(P_y Q_y)}$ for C_{xy} and similarly for other Cs yields Equation 2. The Covariance Theorem for dichotomous attributes was first established by Yule. Paul Lazarsfeld brought it to the attention of American scientists. See "Evidence and Inference in Social Research", Daedalus, 87, 4 (Fall 1958). We have benefited from its illuminating treatment in Hayward Alker, Mathematics and Politics, Macmillan, 1971.

e.g. more Blue firepower, and that the users of contingency planning had more firepower. We still don't have all the relevant data.

It is easy to show that while causation implies correlation, the converse is false: simple correlation does not prove causation. Digging deeper into the data by examining the partial correlations within the two training subgroups between use of contingency planning and success in combat, we find a perfect positive correlation within the subgroup that had West Point training ($\Phi_{xyc} = 1.00$) and a moderate negative correlation within the untrained subgroup ($\Phi_{xy\sim c} = -.50$). So with these data, the relationship between use of contingency planning and success in combat is clearly confounded by training. Indeed, there is an interaction here between use of contingency planning and level of training in impacting combat outcome: with these data, it appears that use of contingency planning caused improved combat effectiveness only under the condition where the commander had West Point training; otherwise, it did not. Hence the findings from partial correlations with these data do not unequivocally support the general hypothesis that use of contingency planning causes successful combat outcomes. So we need a comparison group, more like the treatment group, that is unconfounded by extraneous variables.

We are obliged to consider the role of control variables in general in testing our causal hypotheses. Equation (2) above states the general Covariance/Partial Correlation theorem for correlation of three dichotomous variables, independent (x), dependent (y) and control (c). According to the Theorem, any universal xy correlation is composed of a weighted average of the correlations within control subgroups plus the product of the independent and dependent variable correlations with the control variable. We assume, of course, that independent and control variables precede the dependent variable in time. Control variables, c, which are uncorrelated with the dependent variable, y, are not plausible explanatory factors in the first place; but those that are so correlated may provide rival explanations if they are also correlated with the independent variable, x. Bearing this in mind, is it possible to find a way to conduct an unequivocal test of our causal hypothesis on contingency planning and warfighting effectiveness? In the words of Lazarsfeld [4], " If we have a relationship between x and y and if for any antecedent test factor, c, the partial relationship between x and y does not disappear, then the original relationship should be called a causal one." But do we have a way to examine all plausible test factors?

Controlled Experimentation. It is here that we must advance from the passive Aristotelian mode of empirical investigation to the active, experimental Galilean mode. In the assertions of the Nobel laureate Herb Simon [5], and John Stuart Mill[6] a century earlier, the causal interpretation of a simple (or partial) correlation depends upon the presence of a compatible causal hypothesis and the absence of a plausible rival hypothesis to explain the correlation on other grounds. But Yule's Covariance Theorem,

---

[4] Lazarsfeld, P. Evidence and Inference in Social Research. *Daedalus*. 87, 4, Fall 1958.

[5] Simon, H. A. *Models of Man*, NY: Wiley, 1957.

[6] Cook, T. and Campbell, D. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin, 1979.

(2) above, states that any correlation can be decomposed into the weighted average of the partial correlations within control subgroups plus the product of the independent and dependent variable correlations with the control variable. Hence any new control variable, or combination of control variables, may provide a potential new rival explanation while washing out the original xy correlation in the subtables of partials. Thus in testing our hypothesis that contingency planning causes improved combat effectiveness, we should control not only for training but also for Blue firepower advantage, Blue troop advantage, quality of C4I and other factors. Through the judicious use of control variables, which usually are not completely specified, we could then investigate the persistence of the original xy correlation in the control subtables as significant partial correlations and perhaps prune rival explanatory hypotheses down to a surviving root cause; but beyond successive prunings, the conduct of a controlled experiment enables us to ascertain precisely whether an alleged cause is a real root cause.

In a controlled experiment, the observed subjects (or units) are randomly assigned to the treatment group, here use of contingency planning, x, or to the control group, ~x, and the mean effectiveness of their combat performance, y or ~y, is measured and compared. Since the two groups are now statistically equivalent, any discovered difference in performance between the two groups is due solely to the treatment condition. Just such a procedure is followed in running clinical trials in the modern health sciences to determine the true efficacy of potential new healing drugs.[7] In the biologist's terms, this practice ensures that x is truly an exogenous variable. Ultimately, a controlled experiment affords the best causal test prospect, and it differs from a passive correlational study precisely because the process of active randomization disrupts any lawful relationship between, c, the characteristics of the antecedents of the subjects, e.g. training, and their exposure to the treatment condition, x, i.e. randomization in a controlled experiment effectively sets the value of the correlation between the independent variable (treatment condition) and any control variable to zero, $\phi_{cx} = 0$. Since $\phi_{cx} = 0$ in Equation 2 above for controlled experiments, the universal correlation, ϕ_{xy}, equals simply the weighted average of the partial correlations, ϕ_{xyc} and $\phi_{xy\sim c}$, for all c's: the spurious portion of the xy correlation, $\phi_{yc}\phi_{cx}$, has been nullified. Hence, it follows as in Lazarsfeld's assertion above that, in the context of a controlled experiment, if an observed correlation between x and y is significantly greater than zero, then the hypothesized relationship should be called a

---

[7] Popper, K. R. *The Logic of Scientific Discovery*, NY: Basic Books, 1959. For a generalization of Lazarsfeld's work on causality see H. Simon, "Spurious Correlation: A Causal Interpretation," *op. cit.* pp.42-43. Simon shows that for multivariate causal modeling, using interval scale data, if and only if one can ensure the proper time sequencing of the variables and ensure that the error terms of the variables are uncorrelated with each other, is it safe to assume that the other variables are in fact "controlled for" or "held constant" or correctly "assumed to be random" as required for true causal relationships to be inferred. Our controlled experiment satisfies these conditions since in this context, X is a random variable with a random error term, u_x, and its correlation with u_y is necessarily zero. Otherwise there could exist some prior variable, C, spuriously affecting both X and Y and contributing to both u_x and u_y. For non-experimental investigations involving two, three or more variables, it is necessary carefully to examine the validity of the assumption that the residual error terms of the variables are pairwise uncorrelated with each other. Thus, regardless of whether one's research is experimental or non-experimental, the investigator must somehow isolate sub-systems of variables from the complex environment, verify the non-correlation of residual error terms, and make careful use of controls in order to draw legitimate causal inferences.

causal one. In Simon's terms, there is no tenable rival hypothesis to explain the correlation on other grounds. Thus controlled experiments provide the scientist a probative way of posing causal questions to nature such that her reply will always be revealing and sometimes profound.

Some Controlled Experimental Tests of Causal Hypotheses on Combat Effectiveness

Following this approach, a controlled experimental test of our hypothesis that use of contingency planning causes improved combat effectiveness (H1) was conducted in the Army Training and Doctrine Command Analysis Center (TRAC) Lab at the Naval Postgraduate School (NPS) in January 1988. [8]

H1: Use of contingency planning causes improved combat effectiveness.

This experiment utilized the fine grained, two-sided JANUS wargame simulator (hence the name JANUS for the two-faced Roman god) to provide a realistic combat setting as well as the capability to adjudicate combat moves and measure combat outcome in terms of the summated losses of Red and Blue warfighting platforms over the course of the combat. The validity of the JANUS simulator has been previously tested by comparing the time-course of the Red and Blue attrition data at the battalion level from JANUS-T to comparable data from low intensity laser battles conducted by troops engaged in live exercises at the National Training Center. The fit was found to be "strikingly similar during the force on force part of the battle." [9] Twelve military officers, who were students at NPS, were randomly assigned to one of three, four man teams. Each team played all four possible conditions resulting from crossing contingency planning/ single thread planning with high and low battle intensity. This procedure yielded a total of twelve, three hour trials, half of which involved the use of contingency planning. There were no significant differences between the trials in training, numbers of Red and Blue troops, available firepower, or available C4I. The question was, would use/non-use of contingency planning make a significant difference in combat outcome. In the combat scenario, US forces opposed Soviet forces who were threatening to close down Bandar Abbas and, with it, all Persian Gulf shipping. The US mission was to prevent Soviet forces from going through the Bam Darzin Pass. The results of the experiment confirmed the contingency planning/successful combat outcome hypothesis: Across the sixteen trials, use of contingency planning resulted in a 16 percent advantage to Blue in terms of attrition of Red forces per kilometer of advance ($Y = 26$ cf. 22, $p < .001$).

The foregoing experiment is a replication of an earlier contingency planning experiment which was conducted utilizing the Joint Theatre Level Simulator (JTLS) in the War Lab at NPS in August 1987. [10]   The subjects consisted of two random

---

[8] Needalman, A., Mikaelian, D., Entin, E., and Tenney, R. Contingency Planning in Headquarters*, Proceedings of the JDL BRG C2 Research Symposium*, June 1988.

[9] Ingber, L., Mathematical Comparison of Combat Models to Exercise Data, *Proceedings of the JDL BRG C2 Research Symposium,* June 1989.

[10] MacMillan, J., Entin, E., and Lenz, P. Experiment Report: The Effects of Option Planning and Battle Workload on C2 Effectiveness*. Technical Report, TR-368*, ALPHATECH, Inc., Burlington, MA, Jan. 1988

assignments of 14 officers to one of two teams, each organized into five command cells. Each team participated in four counterbalanced trials of three hours each for a total of four contingency planning trials and four single thread trials. The warfighting scenario here also involved a Persian Gulf mission defending against a Soviet invasion. Here, again, use of contingency planning produced significantly greater Red losses than single thread planning, yielding a 36% advantage for Blue (Y = .84 cf. .62, p = .02). So the general hypothesis that the use of contingency planning causes improved combat effectiveness is once again supported, and this causal relationship is shown to be invariant with respect to the particular wargame simulator or particular officers involved in the experiment. Furthermore, these observations cannot be accounted for with rival explanations of better training, more firepower, more Blue troops or better C4I since these factors were the same in the experimental and control conditions, and teams were randomly assigned to the different conditions.

   Such controlled experiments have been conducted not only to test causal hypotheses regarding the combat effectiveness of particular military doctrines, but also to test causal hypotheses about the combat effectiveness of potential new C4I technologies

H2: Use of a shared COP causes improved combat effectiveness.

A controlled experimental test of the hypothesis that use of a shared Common Operational Picture (COP) causes improved combat performance was conducted in the MIT Research and Engineering Corporation (MITRE) Command Center Engineering Lab in the summer of 1991.[11] This experiment utilized the Navy's Research and Analysis for Systems Engineering (RESA) wargame simulator for an air/sea battle set in the Persian Gulf. Eight experienced Naval officers were recruited from the faculty of the Naval War College and were joined with four retired Admirals to compose four, three man teams. Each team played two COP trials and two control trials for a total of 16, three hour trials, half of which utilized the COP prototype. There were no significant differences between the trials in numbers of Red and Blue troops, available firepower, training or doctrine. The question was, would teams using a cross echelon shared COP fed by both organic and national sensors perform better in combat than a control team with the high commander using only a national sensor fed big picture view and a pair of subordinate ship captains using only local tactical pictures fed by their organic ship sensors. In the combat scenario, US assets are under attack by Red craft, and the Blue team is required to sort through ambiguous information to determine who the attackers are and then take appropriate combat action. The results of the experiment confirmed the shared COP/combat effectiveness hypothesis: Across the 16 trials the ratio of Red losses to Red plus Blue plus Neutral losses was significantly greater when the Blue teams employed the COP (Y = .68 cf. .54, p = .04).

   The foregoing experiment is a replication of the original COP prototype experiment conducted at the Naval Ocean Systems Center (NOSC) RESA Lab in spring '90 utilizing

---

[11] Hiniker, P. and Entin, E. Examining Cognitive Processing in Command Crises: New HEAT Experiments on Shared Battle Graphics and Time Tagging, *Proceedings of the JDL BRG C2 Research Symposium*, July 1992.

the RESA wargame simulator, which was the first time a prototype of a shared COP was subjected to controlled experimental testing [12]. Six experienced Naval officers were recruited in the San Diego area to join three, crisis tempered, retired Admirals to compose three, three man teams. Each team played four, three hour trials , as above, for a total of twelve trials, half of which used the COP prototype. The Persian Gulf scenario was essentially the same as above. Employing the HEAT /OODA Loop Model, we derived two hypotheses on COP effectiveness: H2, as above, use of COP causes improved combat effectiveness; and, as a mechanism for this, H2B, use of COP causes improved situation assessment accuracy, later dubbed "Situation Awareness".[13] The results were inconclusive for H2, but showed substantial support for H2B: When using the prototype COP, Blue teams displayed significantly higher situation awareness, in terms of the proportion of the mission relevant set of warfighting platforms they were able to identify correctly ($Y = .56$ cf. .50, $p = .02$).

---

[12] Hiniker, P. and Entin, E. The Effects of Shared Battle Graphics on Team Performance in Crisis Situations: HEAT Experimental Results. *Proceedings of the JDL BRG C2 Research Symposium*, July 1990.
[13] The HEAT (Headquarters Effectiveness Assessment Tool) Model measures the speed and accuracy of the command decision cycle composed of a sequence of six phases: monitoring, situation assessment, course of action development, outcome prediction, decision, direction of action, and remonitoring. HEAT was initially applied to higher headquarters planning processes. At the tactical level of command decision making, the abbreviated four phases of the similar OODA loop are applied to the decision cycle: Observe, Orient, Decide, Act, Reobserve. It has become an accepted tenet of military doctrine that warfighters should act fast, and inside the decision cycle of the adversary.

Figure 3. Controlled Experimental Tests of Causal Hypotheses on Combat Effectiveness

Site of Experiment	Wargame Simulator Used	Combat Effectiveness Measure (y): for Exp Group(x) / Control(~x)		Number of Trials Run	Significance of Difference*
H1: Use of contingency planning causes improved combat effectiveness.					
NPS TRAC LAB	JANUS	26	22	12	$p < .001$
NPS WAR LAB	JTLS	.84	.62	8	$p = .02$
H2: Use of shared Common Operational Picture causes improved combat effectiveness.					
MITRE CCEL	RESA	.68	.54	16	$p = .04$
NOSC RESA LAB	RESA	-	-	12	n.s.
(H2 B:	RESA	.56**	.50**	12	$p = .02$)
DISA JDEF	RESA	.61	.42	5	$p = .09$
H3: Use of N-KRS replanning aid causes improved combat effectiveness.					
NOSC RESA LAB	RESA	.56	.59	24	n.s.

* All Significance of Difference probabilities are from the F test tables for the ANOVA used in analyzing the experimental results.

**Situation Awareness Measure defined as the proportion of the mission critical set of Red, Blue, and Neutral warfighting platforms correctly identified when comparing the perceptual Situation Assessment Map produced by the subject with the Ground Truth Map produced by the simulator.

Another test of the shared COP/combat effectiveness hypothesis was carried out through another replication of the COP experiment in DISA's new Joint Demonstration and Evaluation Facillity (JDEF) Lab in the summer of '91 utilizing the RESA wargame simulator[14]. Employing the same Persian Gulf scenario and design as above, albeit with a smaller number of trial runs, more support was found for H2: Across the five trials, the ratio of Red losses to Red plus Blue plus Neutral losses was significantly greater when

---

[14] Hiniker, P. HEAT Experiments: Use of the Experimental Method to Evaluate the Effectiveness of HQ C2 Insertions. *Proceedings of the JDL BRG C2 Research Symposium*, pp 323-331, July 1991.

Blue teams employed the COP ($Y = .61$ cf. $.42$, $p = .09$). In sum, three different controlled experiments, conducted in three different laboratory venues, with three different sets of subjects all provided significant support for the hypothesis that use of a shared COP causes improved situation awareness or improved combat effectiveness. In all three experiments across dozens of trials, akin to dozens of small scale military exercises, the observed superior performance of the Blue teams using the shared COP cannot be explained by their use of more troops, more firepower, better doctrine or better training since all these factors were controlled by randomization of subjects in the design of the experiments. The discovered superior combat performance of the Blue teams that was observed and reported here was due solely to their use of a shared Common Operational Picture. All the observational evidence reported here is consistent with the proposition that use of a shared COP causes improved combat effectiveness; there is no tenable rival hypothesis that accounts for these findings.

H3: Use of N-KRS decision aid improves combat effectiveness.

Another C4I technology hypothesized to cause improved combat effectiveness (H3) was experimentally tested in the NOSC RESA Lab using the RESA wargame simulator in spring '90.[15] This technology, the Navy Knowledge-based Replanning System (N-KRS), was a computerized replanning aid designed to produce rapid air tasking orders for carrier based air strike commanders. Six experienced Naval air strike commanders were recruited to play all four conditions of a two-wave Kamchatka Peninsula targeting scenario. Half of these 24 trials involved use of N-KRS. The results showed no significant difference in the proportion of Red targets successfully destroyed ($Y = .56$ cf. $.59$, p = n.s.). Despite the fact that replanning was accomplished significantly faster by the strike commanders when using the automated N-KRS aid, this advantage was offset in the overall command decision cycle by the fact that the experienced strike commanders made significantly less accurate estimates in their projections of target destruction when using the new aid. Thus H3 was not supported by the controlled experimental results, and N-KRS was sent back to the drawing boards for informed modification.

Conclusions

We have produced proof that controlled experiments provide the only unequivocal tests of causal hypotheses. *Post facto* controlled statistical analyses can approach the validity of such controlled experimental tests, but they seldom, if ever, produce unequivocal tests of causal hypotheses. We have also demonstrated that such controlled experiments are feasible and can be conducted in the warfighting area, in particular, with tests of the efficacy of certain military doctrines and certain C4I technologies alleged to improve command decisionmaking. In the process we have produced significant experimental evidence supporting the twin hypotheses that use of contingency planning and use of a shared COP by Blue commanders cause improved combat effectiveness. These replicated, controlled experimental findings permit of no other explanation for the

---

[15] MacMillan, J., and Shaw, J. Experimental Evaluation of a Knowledge-based Air Strike Mission Planning Aid*, Proceedings of the JDL BRG C2 Research Symposium*, June'90.

observations. As summarized in Figure 3, these controlled experimental observations supporting the two hypotheses on the causes of warfighting effectiveness are robust: they were found and replicated in five different experimental venues, employing three different wargame simulators, Army, Navy, and Joint; and they involved more than 50 runs of man-in-the-loop combat trials with five different sets of U. S. Army, Navy, and Air Force officers. British defence scientists interested in even broader combined operations were participants as expert observers and critics in some of these early experiments. The scientifically sound practice of random assignment of subjects to treatment conditions employed here serves to define a controlled experiment and thereby rule out any rival explanations for significant findings. We have also demonstrated the utility of the method of controlled experimentation to delay the acquisition of certain immature prototyped C4I technologies as not significantly effective, while providing important diagnostics for improvement as part of an evolutionary development program. One may, of course, also make informative and useful observations of factors thought to cause improved combat effectiveness by making careful use of quasi-experimental designs where the randomization requirement is relaxed; but then one is obliged to rule out, as well as possible, all plausible rival explanations for the findings by other means. [16]

Both sets of confirmed experimental findings are consistent with the HEAT or OODA Loop Model of command decision making: use of the shared COP makes for more accurate situation awareness, or Observation, among Blue warfighters; use of contingency planning permits of more rapid response, or Orient-Decide-Act time, for a changed situation. Currently, both propositions are also in accord with new DoD emphases in the Defense Transformation: use of the shared COP contributes to "information superiority"; use of contingency planning contributes to "flexible response." Historically, DISA adopted the COP in 1995, converting it from an idea and a prototype into an integral part of the Global Command and Control System (GCCS), now DoD's official C2 system. DISA has evolved and spread the COP continuously since 1995, now to more than 600 sites including the National Military Command Center and all Combatant Commander command centers. Recently the COP has been folded into the Global Information Grid as part of DISA's new Net Centric Enterprises Services. Indeed without a shared COP, current DoD emphases on Network Centric Warfare, as opposed to weapons platform based warfare, would not be feasible for our Joint Forces. [17]

The emerging doctrine of Network Centric Warfare (NCW) has broadened the set of C2 variables believed relevant to combat effectiveness through a sharpened focus on the concept of Shared Situational Awareness for a warfighting team. The NCW doctrine is summarily expressed in the form of four tenets: (1) A robustly networked force improves information sharing; (2) Information sharing and collaboration enhance the quality of information and shared situational awareness; (3) Shared situational awareness enables

[16] Campbell, D. and Stanley, J. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1963

[17] Cebrowski, VADM A. and Garstka, J. Network Centric Warfare: Its Origins and Future, *Proceedings of the Naval Institute*, 124:1, pp. 28-35, 1998.

collaboration and self synchronization, and enhances sustainability and speed of command; (4) These, in turn, dramatically increase mission effectiveness.[18]

  NCW Tenet (1) seems self-evident. We have produced some experimental evidence here that supports Tenet (2). The previously reported controlled experiments show that use of a shared COP by a warfighting team causes increases in Situation Awareness, Shared Situation Awareness and Combat Effectiveness.[19] Regarding Tenets (3) and (4), what still needs to be tested experimentally is the causal mechanism by which enhanced Shared Situational Awareness impacts Mission Effectiveness. With the addition of technology for Collaborative Planning, increased Shared Situational Awareness is variously hypothesized to cause increased Combat Effectiveness by increasing Decision Loop Speed (OODA loop speed), or by increasing (Self)Synchronization (the arrangement of warfighters in time and space), or by increasing Speed of Maneuver. To test these causal hypotheses we don't want to know simply what's the difference between military units employing the COP and Collaboration Technology compared with those without these combined technologies; we need to know what difference does it make on Decision Loop Speed, (Self)Synchronization, and Speed of Maneuver, and ultimately on Combat Effectiveness, if we put in the COP and Collaborative Planning technology compared to what would have happened without the introduction of this technology. While much valuable empirical evidence has been gathered from exercise data and other analytic studies, these causal hypotheses embedded in the tenets of NCW can only be unequivocally tested with controlled experimentation.[20] That such controlled experimentation is feasible is demonstrated by the experimental results reported above.

.

---

[18] DoD Report to Congress on Network Centric Warfare, July 2001.

[19] Hiniker, P. and Entin, E. *The Effects of Shared Battle Graphics on Team Performance in Crisis Situations*, op. cit., 1990 and Hiniker P. and Entin E. *Examining Cognitive Processing in Command Crises*, op. cit., 1992. The first study introduces the concepts of situation awareness and shared situation awareness in the results and conclusions sections.

[20] For a fine assemblage of concepts and extensive data from military exercises bearing directly on NCW see Alberts, D.,Garstka, J., Hayes,R., and Signori, D. *Understanding Information Age Warfare.* Washington, D.C.: DoD CCRP, August 2001.

C3X: Correlation, Causation and Controlled eXperimentation for C2

Dr. Paul J. Hiniker
Sr. Ops Research
September 14, 2004

The Problem

How does one test causal hypotheses on C2 effectiveness against empirical evidence?

Causation

• **All we observe are covariations.**
  (David Hume, 1740)

• **The causal interpretation of a simple(or partial)**
  correlation depends upon

  • **the presence of a compatible causal hypothesis**

  •**and the absence of a plausible rival hypothesis**
   to explain the correlation on other grounds.
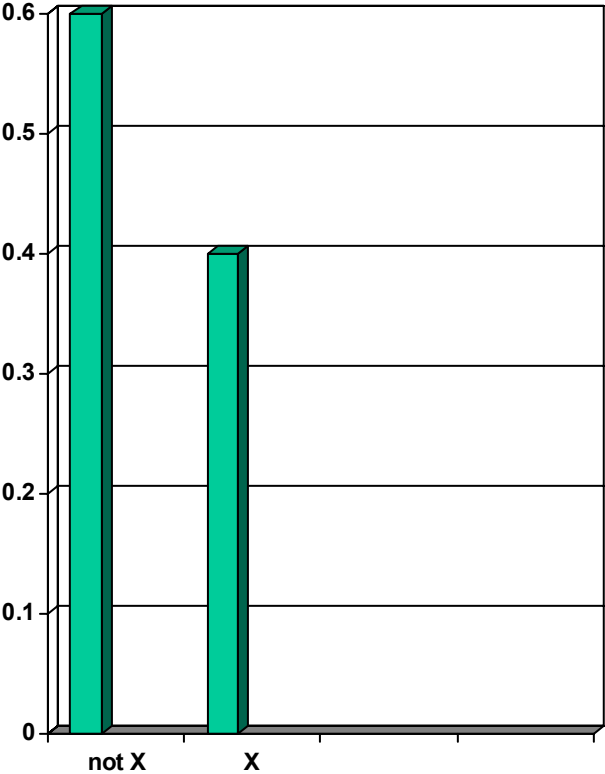
  (Herb Simon, 1957)

Causal Hypotheses & Correlation

Hypothesis: Fire engines prevent fire damage.

X---→Y $Y_X < Y_{not\ X}$ $\Phi_{XY} < 0$

	not X	X	
Y	**.30**	**.20**	.50
not Y	**.20**	**.30**	.50
	.50	.50	N=100

X = Fire engines = Fleet of 4+

Y = Percent of fires w/ damage > $500k(Expected)

40% < 60%

$$\Phi_{XY} = P_{XY} - P_X P_Y / \sqrt{(P_X Q_X P_Y Q_Y)} = -.20$$

X ► Y
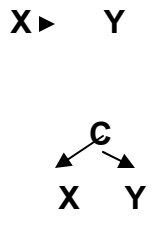
	not X	X	
Y	.2	.3	.50
notY	.3	.2	.50
	.50	.50	100

$\Phi_{xy} = .20$

X = Fire Engines Sent

Y = Fire Damage at Site

C = Size of Fire >1,000 ft

```
    C
   ↙ ↘
  X   Y
```

notC

	notX	X	
Y	.2	.0	.20
notY	.6	.2	.80
	.50	.50	50

$\Phi_{xy-c} = -.25$

C

	notX	X	
Y	.2	.6	.80
notY	.0	.2	.20
	.50	.50	50

$\Phi_{XYC} = -.25$

notC   C

	notC	C	
Y	.1	.4	.50
notY	.4	.1	.50
	.50	.50	100

$\Phi_{YC} = .60$

	notC	C	
X	.1	.4	.50
notX	.4	.1	.50
	.50	.50	100

$\Phi_{XC} = .60$

Yule's (Covariance) Theorem for Dichotomous Attributes

$$\Phi_{XY} =$$

$$\Phi_{XY\text{-}c}P_{\text{-}c}\sqrt{(P_{X\text{-}c}\,Q_{X\text{-}c}P_{Y\text{-}c}Q_{Y\text{-}c}/P_XQ_XP_YQ_Y)} + \Phi_{XYC}P_C\,\sqrt{(P_{XC}Q_{XC}P_{YC}Q_{YC}/P_XQ_XP_YQ_Y)}$$

$$+\,\Phi_{YC}\Phi_{XC}$$

• **For any two attributes, X and Y, and a third control attribute, C , the universal covariance can be decomposed into**
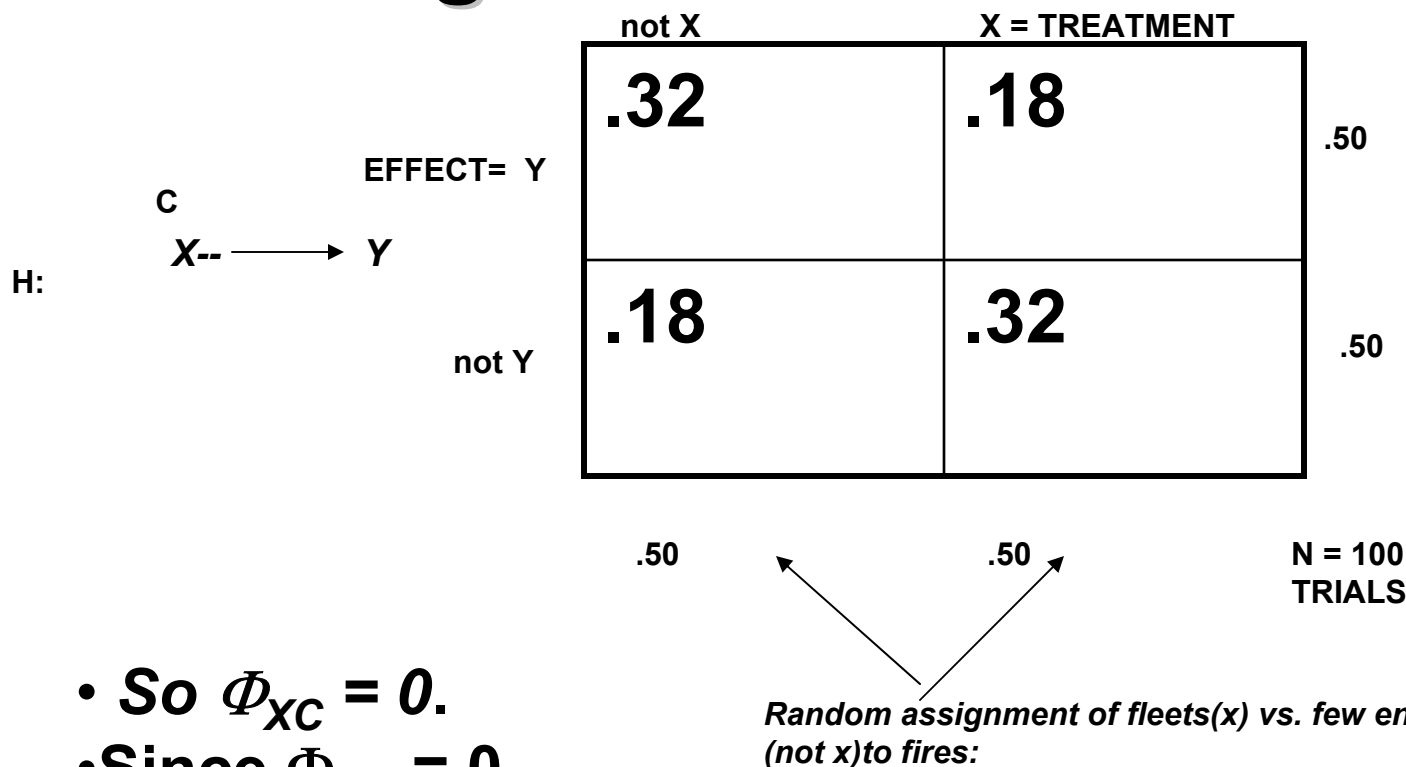
a weighted average of the covariances within control subgroups , and, in addition

a term involving the product of the covariances between Y and C and C and X.

* N.B. In treating causality we assume, of course, that X and C are antecedent to Y.

Controlled Experiment: Fire Engines Prevent Fire Damage

	not X	X = TREATMENT	
EFFECT= Y	**.32**	**.18**	**.50**
not Y	**.18**	**.32**	**.50**
	.50	**.50**	**N = 100 TRIALS**

C

$X-- \longrightarrow Y$

H:

Random assignment of fleets(x) vs. few engines (not x)to fires:

- *So $\Phi_{XC} = 0$.*
- Since $\Phi_{XC} = 0$, Experiment $\Phi_{XY} = (w' \Phi_{XY-C} + w \Phi_{XYC})/ 2 = \Phi_{XY.C}$, for $\forall c$, thus ruling out rival explanations.
- Experiment $\Phi_{XY} = -.28$
- So less fire damage is due to more fire engines on site.

Causal Modeling with Non-Experimental Data

- So to prevent spurious correlation, conduct of a controlled experiment guarantees $\phi_{cx} = 0$ and ensures a valid test of a causal hypothesis.

- However, for non-experimental causal modeling,with one or more independent variables, one must verify that the residual error terms of all the variables are uncorrelated:

  $r_{Uy\,Uxi} = r_{Uxi\,Uxj} = 0$, for all X_i.

  Otherwise, there could exist some extraneous variable(s), C_i, affecting both Y and X_i, hence forming part of u_y and u_{xi}, which would then be correlated; this would spuriously contribute to the correlations implied by the model.
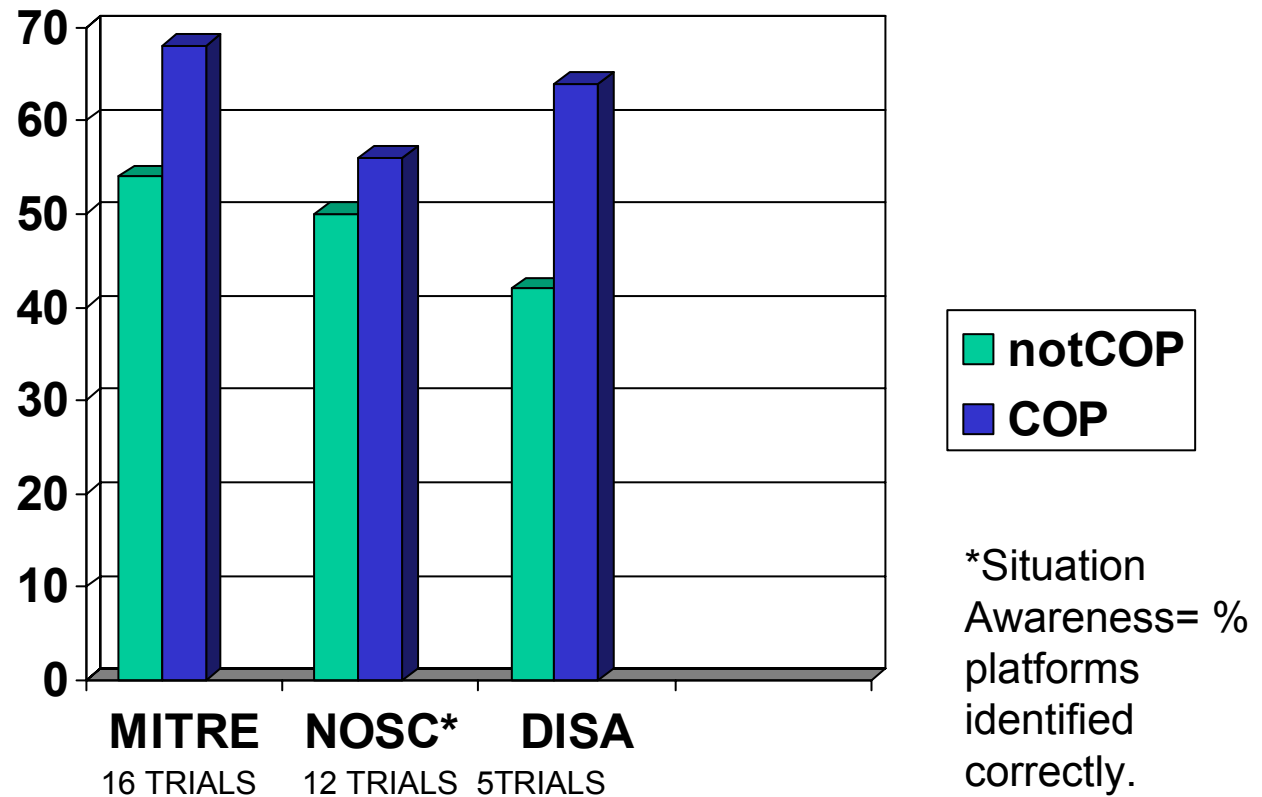
- Thus simple correlation can neither prove nor disprove a causal hypothesis.

Controlled Experiments in C2

H: Use of a shared Common Operational Picture by a combat team(X) causes improved combat effectiveness(Y, in % platforms lost that are Red).

H: X ——→ Y

*Situation Awareness= % platforms identified correctly.

Some Causal Hypotheses on NCW

• A basic assumption underlying most technological acquisitions for defense is the belief that the acquired capability will cause improved military effectiveness; therefore, controlled experimentation should be an integral part of the acquisition process.

• Net Centric Warfare (NCW) doctrine clearly includes such assumptions and several specific causal hypotheses such as the following:

H: Increased Shared Situation Awareness and Collaborative Planning by a distributed combat team causes increased decision loop speed and increased combat effectiveness.

•Such causal hypotheses warrant experimental testing.

Global Net-Centric Solutions -- The Warfighter's Edge